Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation

Michael B. Giles* Lukasz Szpruch†

Abstract

In this paper we introduce a new multilevel Monte Carlo (MLMC) estimator for multidimensional SDEs driven by Brownian motion. Giles has previously shown that if we combine a numerical approximation with strong order of convergence $O(\Delta t)$ with MLMC we can reduce the computational complexity to estimate expected values of functionals of SDE solutions with a root-mean-square error of ϵ from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$. However, in general, to obtain a rate of strong convergence higher than $O(\Delta t^{1/2})$ requires simulation, or approximation, of Lévy areas. In this paper, through the construction of a suitable antithetic multilevel correction estimator, we are able to avoid the simulation of Lévy areas and still achieve an $O(\Delta t^2)$ variance for smooth payoffs, and almost an $O(\Delta t^{3/2})$ variance for piecewise smooth payoffs, even though there is only $O(\Delta t^{1/2})$ strong convergence. This results in an $O(\epsilon^{-2})$ complexity for estimating the value of European and Asian put and call options.

Key words: *Monte Carlo, Multilevel, Lévy area, Stochastic Differential Equation.*

2000 Mathematics Subject Classification: 65C30, 65C05

1 Introduction

In many financial engineering applications, one is interested in the expected value of a financial derivative whose payoff depends upon the solution of a stochastic differential equation (SDE). Using a simple Monte Carlo method with a numerical discretisation with first order weak convergence, to achieve a root-mean-square error of ϵ would require $O(\epsilon^{-2})$ independent paths, each with $O(\epsilon^{-1})$ time steps, giving a computational complexity which is $O(\epsilon^{-3})$, [3].

Recently, Giles [6] introduced a multilevel Monte Carlo (MLMC) estimator which enables a reduction of this computational cost to $O(\epsilon^{-2}(\log \epsilon)^2)$ for Lipschitz payoffs when using the Euler-Maruyama discretisation. Subsequent research using the Milstein discretisation [5] improved this

*Mathematical Institute, University of Oxford, Oxford OX1 3LB, UK (`mike.giles@maths.ox.ac.uk`).
†Mathematical Institute, University of Oxford, Oxford OX1 3LB, UK (`szpruch@maths.ox.ac.uk`).

to $O(\epsilon^{-2})$ for a variety of non-smooth and path-dependent options based on the solution of a scalar SDE. However, a weakness of the Milstein discretisation is that in multiple dimensions it generally requires the simulation of iterated Itô integrals known as Lévy areas, for which there is no known efficient method except in dimension 2 [4, 15, 16].

We consider a general class of multi-dimensional SDEs driven by Brownian motion. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions, and let $w(t)$ be a D-dimensional Brownian motion defined on the probability space. We consider the numerical approximation of SDEs of the form

$$\mathrm{d}x(t) = f(x(t))\ \mathrm{d}t + g(x(t))\ \mathrm{d}w(t), \tag{1.1}$$

where $x(t) \in \mathbb{R}^d$ for each $t \geq 0$, $f \in C^2(\mathbb{R}^d, \mathbb{R}^d)$, $g \in C^2(\mathbb{R}^d, \mathbb{R}^{d\times D})$, and for simplicity we assume a fixed initial value $x_0 \in \mathbb{R}^d$.

In this paper we are primarily concerned with estimating $\mathbb{E}[P(x(T))]$, the expected value of a payoff depending on the solution at a fixed time T, Defining the tensor $h_{ijk}(x)$ as

$$h_{ijk}(x) = \tfrac{1}{2} \sum_{l=1}^{d} g_{lk}(x)\ \frac{\partial g_{ij}}{\partial x_l}(x), \tag{1.2}$$

when using N uniform timesteps $\Delta t = T/N$, the i^{th} component of the first order Milstein approximation $X_n \approx x(n\,\Delta t)$ has the form [12]

$$\hat{X}_{i,n+1} = \hat{X}_{i,n} + f_i(\hat{X}_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(\hat{X}_n)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(\hat{X}_n)\,(\Delta w_{j,n}\Delta w_{k,n} - \Omega_{jk}\,\Delta t - A_{jk,n}) \tag{1.3}$$

where Ω is the correlation matrix for the driving Brownian paths, and $A_{jk,n}$ is the Lévy area defined as

$$A_{jk,n} = \int_{t_n}^{t_{n+1}} \big(w_j(t) - w_j(t_n)\big)\ \mathrm{d}w_k(t) - \int_{t_n}^{t_{n+1}} \big(w_k(t) - w_k(t_n)\big)\ \mathrm{d}w_j(t).$$

In some applications, the diffusion coefficient $g(x)$ has a commutativity property which gives $h_{ijk}(x) = h_{ikj}(x)$ for all i, j, k. In that case, because the Lévy areas are anti-symmetric (i.e. $A_{jk,n} = -A_{kj,n}$), it follows that $h_{ijk}(X_n)\,A_{jk,n} + h_{ikj}(X_n)\,A_{kj,n} = 0$ and therefore the terms involving the Lévy areas cancel and so it is not necessary to simulate them. However, this only happens in special cases.

Clark & Cameron [2] proved for a particular SDE that it is impossible to achieve a better order of strong convergence than the Euler-Maruyama discretisation when using just the discrete increments of the underlying Brownian motion. The analysis was extended by Müller-Gronbach [14] to general SDEs. As a consequence if we use the standard MLMC method with the Milstein scheme without simulating the Lévy areas the complexity will remain the same as for Euler-Maruyama. Nevertheless, in this paper we show that by constructing a suitable antithetic estimator one can neglect the Lévy areas and still obtain a multilevel correction estimator with a variance which decays at the same rate as the scalar Milstein estimator.

We begin the paper by reviewing the multilevel Monte Carlo approach, introducing the idea of the antithetic estimator and bounding the behaviour of its variance under certain conditions. Because of its simplicity, we then consider Clark & Cameron's model problem, and prove that the antithetic path simulations do satisfy the required conditions to give an $O(\Delta t^2)$ variance

convergence for a smooth payoff. This then motivates the subsequent analysis for the general class of multidimensional SDEs. The appendix contains the detailed proofs of the key theorems.

In this paper we restrict attention to financial applications with either a European payoff, dependent on the final value $x(T)$, or an Asian payoff, dependent on the average of $x(t)$ over the time interval $[0, T]$. It is proved that when the payoff is twice differentiable, with bounded derivatives, the rate of convergence of the multilevel correction variance is doubled from $O(\Delta t)$ to $O(\Delta t^2)$. If the payoff is Lipschitz, and twice differentiable almost everywhere, then the rate of convergence is reduced to $O(\Delta t^{3/2})$, but this is still sufficient to make the overall complexity $O(\epsilon^{-2})$ to achieve a root-mean-square accuracy of ϵ.

2 Multilevel Monte Carlo estimation

2.1 MLMC estimators

In its most general form, multilevel Monte Carlo simulation uses a number of levels of resolution, $\ell = 0, 1, \ldots, L$, with $\ell = 0$ being the coarsest, and $\ell = L$ being the finest. In the context of a SDEs simulation, level 0 may have just one timestep for the whole time interval $[0, T]$, whereas level L might have 2^L uniform timesteps.

If P denotes the payoff (or other output functional of interest), and P_ℓ denote its approximation on level l, then the expected value $\mathbb{E}[P_L]$ on the finest level is equal to the expected value $\mathbb{E}[P_0]$ on the coarsest level plus a sum of corrections which give the difference in expectation between simulations on successive levels,

$$\mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}]. \tag{2.1}$$

The idea behind MLMC is to independently estimate each of the expectations on the right-hand side of (2.1) in a way which minimises the overall variance for a given computational cost. Let Y_0 be an estimator for $\mathbb{E}[P_0]$ using N_0 samples, and let Y_ℓ, $\ell > 0$, be an estimator for $\mathbb{E}[P_\ell - P_{\ell-1}]$ using N_ℓ samples. The simplest estimator is a mean of N_ℓ independent samples, which for $\ell > 0$ is

$$Y_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} (P_\ell^i - P_{\ell-1}^i). \tag{2.2}$$

The key point here is that $P_\ell^i - P_{\ell-1}^i$ should come from two discrete approximations for the same underlying stochastic sample, so that on finer levels of resolution the difference is small (due to strong convergence) and so the variance is also small. Hence very few samples will be required on finer levels to accurately estimate the expected value.

Here we recall the Theorem from [8] (which is a slight generalisation of the original theorem in [6]) which gives the complexity of MLMC estimation.

Theorem 2.1. *Let P denote a functional of the solution of a stochastic differential equation, and let P_ℓ denote the corresponding level ℓ numerical approximation. If there exist independent estimators Y_ℓ based on N_ℓ Monte Carlo samples, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2} \min(\beta, \gamma)$ and*

i) $|\mathbb{E}[P_\ell - P]| \leq c_1\, 2^{-\alpha\, \ell}$

ii) $\mathbb{E}[Y_\ell] = \begin{cases} \mathbb{E}[P_0], & \ell = 0 \\ \mathbb{E}[P_\ell - P_{\ell-1}], & \ell > 0 \end{cases}$

iii) $\mathbb{V}[Y_\ell] \leq c_2\, N_\ell^{-1} 2^{-\beta\, \ell}$

iv) $C_\ell \leq c_3\, N_\ell\, 2^{\gamma\, \ell}$, *where* C_ℓ *is the computational complexity of* Y_ℓ

then there exists a positive constant c_4 *such that for any* $\epsilon < e^{-1}$ *there are values* L *and* N_ℓ *for which the multilevel estimator*

$$Y = \sum_{\ell=0}^{L} Y_\ell,$$

has a mean-square-error with bound

$$MSE \equiv \mathbb{E}\left[\left(Y - \mathbb{E}[P] \right)^2 \right] < \epsilon^2$$

with a computational complexity C *with bound*

$$C \leq \begin{cases} c_4\, \epsilon^{-2}, & \beta > \gamma, \\ c_4\, \epsilon^{-2} (\log \epsilon)^2, & \beta = \gamma, \\ c_4\, \epsilon^{-2-(\gamma-\beta)/\alpha}, & 0 < \beta < \gamma. \end{cases}$$

In (2.2) we have used the same estimator for the payoff P_ℓ on every level ℓ, and therefore (2.1) is a trivial identity due to the telescoping summation. However, in [5] Giles numerically showed that it can be better to use different estimators for the finer and coarser of the two levels being considered, P_ℓ^f when level ℓ is the finer level, and P_ℓ^c when level ℓ is the coarser level. In this case, we require that

$$\mathbb{E}[P_\ell^f] = \mathbb{E}[P_\ell^c] \quad \text{for } \ell = 1, \ldots, L, \tag{2.3}$$

so that

$$E[P_L^f] = \mathbb{E}[P_0^f] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell^f - P_{\ell-1}^c].$$

The MLMC Theorem is still applicable to this modified estimator. The advantage is that it gives the flexibility to construct approximations for which $P_\ell^f - P_{\ell-1}^c$ is much smaller than the original $P_\ell - P_{\ell-1}$, giving a larger value for β, the rate of variance convergence in condition *iii)* in the theorem.

2.2 Antithetic MLMC estimator

Based on the well-known method of antithetic variates (see for example [10]), the idea for the antithetic estimator is to exploit the flexibility of the more general MLMC estimator by defining $P_{\ell-1}^c$ to be the usual payoff $P(X^c)$ coming from a level $\ell - 1$ coarse simulation X^c, and define

P_ℓ^f to be the average of the payoffs $P(X^f), P(X^a)$ coming from an antithetic pair of level ℓ simulations, X^f and X^a.

X^f will be defined in a way which corresponds naturally to the construction of X^c. Its antithetic "twin" X^a will be defined so that it has exactly the same distribution as X^f, conditional on X^c, which ensures that $\mathbb{E}[P(X^f)] = \mathbb{E}[P(X^a)]$ and hence (2.3) is satisfied, but at the same time

$$\left(X^f - X^c\right) \;\approx\; -\left(X^a - X^c\right)$$

and therefore

$$\left(P(X^f) - P(X^c)\right) \;\approx\; -\left(P(X^a) - P(X^c)\right),$$

so that $\frac{1}{2}\left(P(X^f) + P(X^a)\right) \approx P(X^c)$. This leads to $\frac{1}{2}\left(P(X^f) + P(X^a)\right) - P(X^c)$ having a much smaller variance than the standard estimator $P(X^f) - P(X^c)$.

We now present a lemma which motivates the rest of the paper by giving an upper bound on the convergence of the variance of $\frac{1}{2}\left(P(X^f) + P(X^a)\right) - P(X^c)$.

Lemma 2.2. *If $P \in C^2(\mathbb{R}^d, \mathbb{R})$ and there exist constants L_1, L_2 such that for all $x \in \mathbb{R}^d$*

$$\left\| \frac{\partial P}{\partial x} \right\| \leq L_1, \qquad \left\| \frac{\partial^2 P}{\partial x^2} \right\| \leq L_2.$$

then for $p \geq 2$,

$$\begin{aligned}
&\mathbb{E}\left[\left(\tfrac{1}{2}(P(X^f) + P(X^a)) - P(X^c) \right)^p \right] \\
&\qquad \leq\; 2^{p-1}\, L_1^p\; \mathbb{E}\left[\left\| \tfrac{1}{2}(X^f + X^a) - X^c \right\|^p \right] \;+\; 2^{-(p+1)}\, L_2^p\; \mathbb{E}\left[\left\| X^f - X^a \right\|^{2p} \right].
\end{aligned}$$

Proof. If we define $\overline{X}^f \equiv \frac{1}{2}(X^f + X^a)$, then a Taylor expansion gives

$$P(X^f) = P(\overline{X}^f) + \frac{\partial P}{\partial x}^T(\overline{X}^f)\; (X^f - \overline{X}^f) + \tfrac{1}{2}(X^f - \overline{X}^f)^T\, \frac{\partial^2 P}{\partial x^2}(\xi_1)\; (X^f - \overline{X}^f)$$

for some ξ_1 on the line between $\overline{X}^f$ and X^f. Performing a similar expansion for $P(X^a)$ and then averaging the two, the linear terms cancel and one obtains

$$\begin{aligned}
\tfrac{1}{2}(P(X^f) + P(X^a)) \;&=\; P(\overline{X}^f) + \tfrac{1}{4}(X^f - \overline{X}^f)^T\, \frac{\partial^2 P}{\partial x^2}(\xi_1)\; (X^f - \overline{X}^f) \\
&\qquad + \tfrac{1}{4}(X^a - \overline{X}^f)^T\, \frac{\partial^2 P}{\partial x^2}(\xi_2)\; (X^a - \overline{X}^f) \\
&=\; P(\overline{X}^f) + \tfrac{1}{8}(X^f - X^a)^T\, \frac{\partial^2 P}{\partial x^2}(\xi_3)\; (X^f - X^a)
\end{aligned}$$

for some ξ_3 on the line between X^a and X^f, due to the mean value theorem. We then obtain

$$\tfrac{1}{2}(P(X^f) + P(X^a)) - P(X^c) = \frac{\partial P}{\partial x}^T(\xi_4)\; (\overline{X}^f - X^c) + \;\tfrac{1}{8}(X^f - X^a)^T\, \frac{\partial^2 P}{\partial x^2}(\xi_3)\; (X^f - X^a),$$

for some ξ_4 on the line between $\overline{X}^f$ and X^c. Hence,

$$\left| \tfrac{1}{2}(P(X^f) + P(X^a)) - P(X^c) \right| \;\leq\; L_1 \left\| \overline{X}^f - X^c \right\| \;+\; \tfrac{1}{4}\, L_2 \left\| X^f - X^a \right\|^2,$$

and the final result follows from the standard inequality

$$\left| \sum_{n=1}^{N} a_n \right|^p \leq N^{p-1} \sum_{n=1}^{N} |a_n|^p, \tag{2.4}$$

and then taking the expectation. □

In the multidimensional SDE applications considered in this paper, we will show that the Milstein approximation with the Lévy areas set to zero, combined with the antithetic construction, leads to $X^f - X^a = O(\Delta t^{1/2})$ but $\overline{X}^f - X^c = O(\Delta t)$. Hence, the variance $\mathbb{V}[\frac{1}{2}(P_l^f + P_l^a) - P_{l-1}^c]$ is $O(\Delta t^2)$, which is the order obtained for scalar SDEs using the Milstein discretisation with its first order strong convergence. We first show this for the simple Clark & Cameron model problem which can be analysed in detail. We then extend the analysis to a general class of multidimensional SDEs.

3 Clark-Cameron Example

3.1 Clark-Cameron analysis

The paper of Clark and Cameron [2] addresses the question of how accurately one can approximate the solution of an SDE driven by an underlying multi-dimensional Brownian motion, using only uniformly-spaced discrete Brownian increments. Their model problem is

$$\begin{aligned} \mathrm{d}x_1(t) &= \mathrm{d}w_1(t) \\ \mathrm{d}x_2(t) &= x_1(t)\, \mathrm{d}w_2(t), \end{aligned} \tag{3.1}$$

with $x(0) = y(0) = 0$, and zero correlation between the two Brownian motions $w_1(t)$ and $w_2(t)$. These equations can be integrated exactly over a time interval $[t_n, t_{n+1}]$, where $t_n = n\,\Delta t$, to give

$$\begin{aligned} x_1(t_{n+1}) &= x_1(t_n) + \Delta w_{1,n} \\ x_2(t_{n+1}) &= x_2(t_n) + x_1(t_n)\Delta w_{2,n} + \tfrac{1}{2}\Delta w_{1,n}\Delta w_{2,n} + \tfrac{1}{2}A_{12,n} \end{aligned} \tag{3.2}$$

where $\Delta w_{i,n} \equiv w_i(t_{n+1}) - w_i(t_n)$, and $A_{12,n}$ is the Lévy area defined as

$$A_{12,n} = \int_{t_n}^{t_{n+1}} \big(w_1(t) - w_1(t_n)\big)\, \mathrm{d}w_2(t) - \int_{t_n}^{t_{n+1}} \big(w_2(t) - w_2(t_n)\big)\, \mathrm{d}w_1(t).$$

This corresponds exactly to the Milstein discretisation presented in (1.3), so for this simple model problem the Milstein discretisation is exact.

The point of Clark and Cameron's paper is that for a given set of discrete Brownian increments, the value for $x_1(t_n)$ is determined exactly for all n, but the value for $x_2(t_n)$ depends on the unknown Lévy areas. Since $\mathbb{E}[A_{12,n} \,|\, \Delta w_{1,n}, \Delta w_{2,n}] = 0$, the conditional expected value is given by (3.2) with the Lévy areas set to zero. In addition, it follows that for *any* numerical

approximation $X(T)$ based solely on the set of discrete Brownian increments Δw,

$$\begin{aligned}\mathbb{E}[(x_2(T)-X_2(T))^2] &= \mathbb{E}[\ \mathbb{E}[(x_2(T)-X_2(T))^2 \mid \Delta w]\ ] \\ &\geq \mathbb{E}[\ \mathbb{V}[x_2(T) \mid \Delta w]\ ] \\ &= \tfrac{1}{4}\sum_{n=0}^{N-1} \mathbb{V}[A_{12,n}\ ] \\ &= \tfrac{1}{4}\, T\, \Delta t.\end{aligned}$$

Hence, one cannot achieve better than $O(\Delta t^{1/2})$ strong convergence, and the mean square error is minimised when the inequality in the above equation is an equality, which is when

$$X_2(T) = \mathbb{E}[x_2(T) \,|\, \Delta w],$$

which is achieved by setting the Lévy areas set to zero.

3.2 Antithetic MLMC estimator

We define a coarse path approximation X^c with timestep Δt by neglecting the Lévy area terms to give

$$\begin{aligned}X^c_{1,n+1} &= X^c_{1,n} + \Delta w_{1,n} \\ X^c_{2,n+1} &= X^c_{2,n} + X^c_{1,n}\Delta w_{2,n} + \tfrac{1}{2}\,\Delta w_{1,n}\,\Delta w_{2,n}\end{aligned} \tag{3.3}$$

This is equivalent to replacing the true Brownian path by a piecewise linear approximation as illustrated in Figure 1.

Similarly, we define the corresponding two half-timesteps of the first fine path approximation

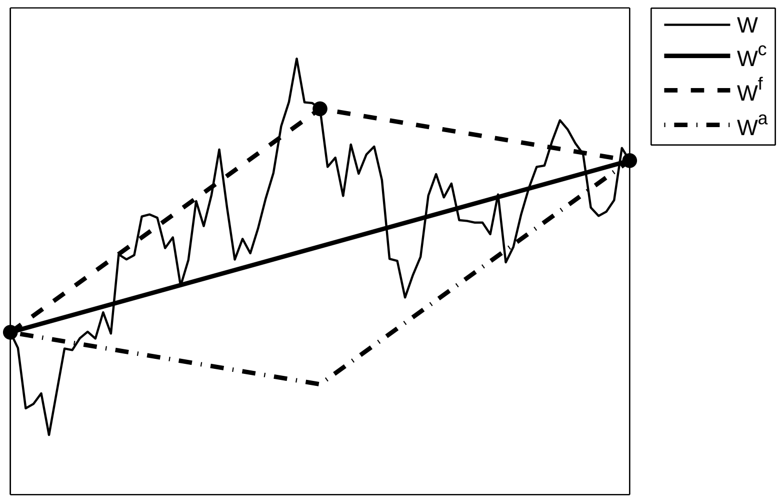

Figure 1: Brownian path and approximations over one coarse timestep

X^f by

$$\begin{aligned}
X^f_{1,n+\frac{1}{2}} &= X^f_{1,n} + \delta w_{1,n} \\
X^f_{2,n+\frac{1}{2}} &= X^f_{2,n} + X^f_{1,n}\,\delta w_{2,n} + \tfrac{1}{2}\,\delta w_{1,n}\,\delta w_{2,n} \\
X^f_{1,n+1} &= X^f_{1,n+\frac{1}{2}} + \delta w_{1,n+\frac{1}{2}} \\
X^f_{2,n+1} &= X^f_{2,n+\frac{1}{2}} + X^f_{1,n+\frac{1}{2}}\,\delta w_{2,n+\frac{1}{2}} + \tfrac{1}{2}\,\delta w_{1,n+\frac{1}{2}}\,\delta w_{2,n+\frac{1}{2}}
\end{aligned}$$

in which $\delta w_n \equiv w(t_{n+\frac{1}{2}}) - w(t_n)$, $\delta w_{n+\frac{1}{2}} \equiv w(t_{n+1}) - w(t_{n+\frac{1}{2}})$ are the Brownian increments over the first and second halves of the coarse timestep, and so $\Delta w_n = \delta w_n + \delta w_{n+\frac{1}{2}}$. Using this relation, the equations for the two fine timesteps can be combined to give an equation for the increment over the coarse timestep,

$$\begin{aligned}
X^f_{1,n+1} &= X^f_{1,n} + \Delta w_{1,n} \\
X^f_{2,n+1} &= X^f_{2,n} + X^f_{1,n}\,\Delta w_{2,n} + \tfrac{1}{2}\,\Delta w_{1,n}\,\Delta w_{2,n} \\
&\quad + \tfrac{1}{2}\left(\delta w_{1,n}\,\delta w_{2,n+\frac{1}{2}} - \delta w_{2,n}\,\delta w_{1,n+\frac{1}{2}}\right).
\end{aligned} \tag{3.4}$$

The antithetic approximation X^a_n is defined by exactly the same discretisation except that the Brownian increments δw_n and $\delta w_{n+\frac{1}{2}}$ are swapped, as illustrated in Figure 1. This gives

$$\begin{aligned}
X^a_{1,n+\frac{1}{2}} &= X^a_{1,n} + \delta w_{1,n+\frac{1}{2}}, \\
X^a_{2,n+\frac{1}{2}} &= X^a_{2,n} + X^a_{1,n}\,\delta w_{2,n+\frac{1}{2}} + \tfrac{1}{2}\,\delta w_{1,n+\frac{1}{2}}\delta\, w_{2,n+\frac{1}{2}}, \\
X^a_{1,n+1} &= X^a_{1,n+\frac{1}{2}} + \delta w_{1,n}, \\
X^a_{2,n+1} &= X^a_{2,n+\frac{1}{2}} + X^a_{1,n+\frac{1}{2}}\,\delta w_{2,n} + \tfrac{1}{2}\,\delta w_{1,n}\,\delta w_{2,n},
\end{aligned}$$

and hence

$$\begin{aligned}
X^a_{1,n+1} &= X^a_{1,n} + \Delta w_{1,n}, \\
X^a_{2,n+1} &= X^a_{2,n} + X^a_{1,n}\,\Delta w_{2,n} + \tfrac{1}{2}\,\Delta w_{1,n}\,\Delta w_{2,n} \\
&\quad - \tfrac{1}{2}\left(\delta w_{1,n}\,\delta w_{2,n+\frac{1}{2}} - \delta w_{2,n}\,\delta w_{1,n+\frac{1}{2}}\right).
\end{aligned} \tag{3.5}$$

Swapping δw_n and $\delta w_{n+\frac{1}{2}}$ does not change the distribution of the driving Brownian increments, and hence X^a has exactly the same distribution as X^f. Note also the change in sign in the last term in (3.4) compared to the corresponding term in (3.5). This is important because these two terms cancel when the two equations are averaged.

These last terms correspond to the Lévy areas for the fine path and the antithetic path, and the sign reversal is a particular instance of a more general result for time-reversed Brownian motion, [11]. If $(w_t, 0 \le t \le 1)$ denotes a Brownian motion on the time interval $[0,1]$ then the time-reversed Brownian motion $(z_t, 0 \le t \le 1)$ defined by

$$z_t = w_1 - w_{1-t}, \tag{3.6}$$

has exactly the same distribution, and it can be shown that its Lévy area is equal in magnitude and opposite in sign to that of w_t.

Lemma 3.1. *If X^f_n, X^a_n and X^c_n are as defined above, then*

$$X^f_{1,n} = X^a_{1,n} = X^c_{1,n}, \qquad \tfrac{1}{2}\left(X^f_{2,n} + X^a_{2,n}\right) = X^c_{2,n}, \qquad \forall n \leq N$$

and

$$\mathbb{E}\left[\left(X^f_{2,N} - X^a_{2,N}\right)^4\right] = \tfrac{3}{4}\, T\,(T{+}\Delta t)\,\Delta t^2.$$

Proof. Comparing (3.3), (3.4) and (3.5), it is clear that $X^f_{1,n}$, $X^a_{1,n}$ and $X^c_{1,n}$ all satisfy the same difference equation and so are equal. Given this, averaging the equations for $X^f_{2,n}$ and $X^a_{2,n}$ gives the same difference equation as for $X^c_{2,n}$, and so therefore $\frac{1}{2}\left(X^f_{2,n} + X^a_{2,n}\right) = X^c_{2,n}$. Finally, summing the difference of the equations for $X^f_{2,n}$ and $X^a_{2,n}$ gives

$$X^f_{2,N} - X^a_{2,N} = \sum_{n=0}^{N-1} \left(\delta w_{1,n}\, \delta w_{2,n+\frac{1}{2}} - \delta w_{2,n}\, \delta w_{1,n+\frac{1}{2}} \right).$$

Since the $\delta w_{j,n}$ are all i.i.d. Normal variables with variance $\frac{1}{2}\Delta t$, it is easily shown that

$$\begin{aligned}
\mathbb{E}[(\delta w_{1,n}\, \delta w_{2,n+\frac{1}{2}} - \delta w_{2,n}\, \delta w_{1,n+\frac{1}{2}})^2] &= \tfrac{1}{2}\,\Delta t^2,\\
\mathbb{E}[(\delta w_{1,n}\, \delta w_{2,n+\frac{1}{2}} - \delta w_{2,n}\, \delta w_{1,n+\frac{1}{2}})^4] &= \tfrac{3}{2}\,\Delta t^4,
\end{aligned}$$

and it then follows that

$$\mathbb{E}[(X^f_{2,N} - X^a_{2,N})^4] \;=\; \left(\tfrac{1}{2}\,\Delta t^2\right)^2\, \frac{N(N{-}1)}{2}\, \frac{4\times 3}{2} + \tfrac{3}{2}\,\Delta t^4\, N \;=\; \tfrac{3}{4}\, T\,(T{+}\Delta t)\,\Delta t^2.$$

In the above derivation, when expanding $(X^f_{2,N} {-} X^a_{2,N})^4$, the first contribution comes from terms of the form $(\delta w_{1,m}\, \delta w_{2,m+\frac{1}{2}} - \delta w_{2,m}\, \delta w_{1,m+\frac{1}{2}})^2 (\delta w_{1,n}\, \delta w_{2,n+\frac{1}{2}} - \delta w_{2,n}\, \delta w_{1,n+\frac{1}{2}})^2$ for $m \neq n$, while the second contribution comes from terms of the form $(\delta w_{1,n}\, \delta w_{2,n+\frac{1}{2}} - \delta w_{2,n}\, \delta w_{1,n+\frac{1}{2}})^4$. All other terms have zero expectation. □

Combining the above result with Lemma 2.2 for $p = 2$ gives a second order bound on the multilevel estimator variance for payoffs satisfying the required smoothness conditions.

4 General theory

4.1 Milstein discretisation

In this section we extend the analysis of the Clark-Cameron example to general the multidimensional SDE (1.1). We make the standard assumptions that f, g and h have a uniform Lipschitz bound, and so have uniformly bounded first derivatives. In addition, we make the assumption that f and g have uniformly bounded second derivatives. More formally

Assumption 4.1. *Let* $f \in C^2(\mathbb{R}^d, \mathbb{R}^d)$ *and* $g \in C^2(\mathbb{R}^d, \mathbb{R}^{d\times m})$. *There exists a constant* L *such that for any* $x \in \mathbb{R}^d$, *and for all* $1 \leq i \leq d$ *and* $1 \leq j,k,l \leq m$

$$\left|\frac{\partial f_i}{\partial x_l}(x)\right| \leq L, \quad \left|\frac{\partial g_{ij}}{\partial x_l}(x)\right| \leq L, \quad \left|\frac{\partial h_{ijk}}{\partial x_l}(x)\right| \leq L, \quad \left|\frac{\partial^2 f_i}{\partial x_k \partial x_l}(x)\right| \leq L, \quad \left|\frac{\partial^2 g_{ij}}{\partial x_k \partial x_l}(x)\right| \leq L,$$

Let us recall that the general Milstein scheme [12] has the form

$$\widehat{X}_{i,n+1} = \widehat{X}_{i,n} + f_i(\widehat{X}_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(\widehat{X}_n)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(\widehat{X}_n)\,(\Delta w_{j,n}\,\Delta w_{k,n} - \Omega_{jk}\Delta t - A_{jk,n})\,. \tag{4.1}$$

As in the Clark-Cameron example, we drop the Lévy areas terms, and instead use the truncated Milstein approximation

$$X_{i,n+1} = X_{i,n} + f_i(X_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(X_n)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(X_n)\,(\Delta w_{j,n}\,\Delta w_{k,n} - \Omega_{jk}\Delta t)\,. \tag{4.2}$$

Under Assumption 4.1 it is a standard result that the moments of the general Milstein approximation X_n are bounded, and X_n strongly converges to the solution of the SDE (1.1); this remains true for the truncated Milstein approximation as stated in the following Lemma.

Lemma 4.2. *For* $p \geq 2$, *there exists a constant* K_p, *independent of the time step, such that*

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|X_n\|^p\right] \leq K_p,$$

and

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|X_n - x(t_n)\|^p\right] \leq K_p\,\Delta t^{p/2}.$$

Proof. The proof in [14] follows the standard method of analysis in references such as [12, 13]. □

Hence, the rate of strong convergence is $O(\Delta t^{1/2})$, which is no better than the Euler-Maruyama discretisation. Nevertheless, we will show that the antithetic multilevel estimator has a variance which converges to zero at the same rate as the full Milstein approximation.

Corollary 4.3. *For* $p \geq 2$, *there exists a constant* K_p, *independent of the time step, such that*

$$\mathbb{E}\left[\max_{0\leq n\leq N} |f_i(X_n)|^p\right] \leq K_p, \quad \mathbb{E}\left[\max_{0\leq n\leq N} |g_{ij}(X_n)|^p\right] \leq K_p, \quad \mathbb{E}\left[\max_{0\leq n\leq N} |h_{ijk}(X_n)|^p\right] \leq K_p,$$

for all $1 \leq i \leq d$ *and* $1 \leq j,k \leq D$.

Proof. The bounded first derivatives of $f(x), g(x), h(x)$ imply that they grow no faster than linearly as $\|x\| \to \infty$, and the result then follows from the bound in Lemma 4.2. □

In order to derive appropriate bounds on the antithetic estimator we also need the following lemma.

Lemma 4.4. *For $p \geq 2$, there exists a constant K_p, independent of the time step, such that*

$$\max_{0\leq n\leq N} \mathbb{E}\left[\,\|X_{n+1}-X_n\|^p\right] \leq K_p\,\Delta t^{p/2}.$$

Proof. We start from (4.2) and inequality (2.4) which gives

$$\begin{aligned}\mathbb{E}\left[|X_{i,n+1}-X_{i,n}|^p\right] \;\leq\; 3^{p-1} \Bigg(& \mathbb{E}\left[|f_i(X_n)\,\Delta t|^p\right] + \mathbb{E}\left[\left|\sum_{j=1}^{D} g_{ij}(X_n)\,\Delta w_{j,n}\right|^p\right] \\ & + \mathbb{E}\left[\left|\sum_{j,k=1}^{D} h_{ijk}(X_n)\;(\Delta w_{j,n}\Delta w_{k,n}-\Omega_{jk}\Delta t)\right|^p\right]\Bigg).\end{aligned}$$

The first term on the right has a $O(\Delta t^p)$ bound due to the uniform bound on $\mathbb{E}\left[|f_i(X_n)|^p\right]$. For the second term we note that because $\Delta w_{j,n}$ is independent of X_n then

$$\mathbb{E}\left[\left|\sum_{j=1}^{D} g_{ij}(X_n)\,\Delta w_{j,n}\right|^p\right] \leq D^{p-1}\sum_{j=1}^{D}\mathbb{E}\left[\,|g_{ij}(X_n)|^p\right]\;\mathbb{E}\left[\,|\Delta w_{j,n}|^p\right],$$

and we obtain a $O(\Delta t^{p/2})$ bound due to the uniform bound on $\mathbb{E}\left[|g_{ij}(X_n)|^p\right]$ and standard results for the moments of Brownian increments. The third term is handled in a similar way and has a $O(\Delta t^p)$ bound.

Together these give a $O(\Delta t^{p/2})$ bound for $\mathbb{E}\left[\,|X_{i,n+1}-X_{i,n}|^p\right]$ for each i, and hence also for $\mathbb{E}\left[\,\|X_{n+1}-X_n\|^p\right]$. □

4.2 Antithetic MLMC estimator

Using the coarse timestep Δt, the coarse path approximation X^c_n, is given by the Milstein approximation without the Lévy area term,

$$X^c_{i,n+1} = X^c_{i,n} + f_i(X^c_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(X^c_n)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(X^c_n)\,(\Delta w_{j,n}\,\Delta w_{k,n} - \Omega_{jk}\,\Delta t)\,.$$

The first fine path approximation X^f_n uses the corresponding discretisation with timestep $\Delta t/2$,

$$\begin{aligned} X^f_{i,n+\frac12} &= X^f_{i,n} + f_i(X^f_n)\,\Delta t/2 + \sum_{j=1}^{D} g_{ij}(X^f_n)\,\delta w_{j,n} \\ &\quad + \sum_{j,k=1}^{D} h_{ijk}(X^f_n)\,(\delta w_{j,n}\,\delta w_{k,n} - \Omega_{jk}\,\Delta t/2)\,, \end{aligned} \tag{4.3}$$

$$\begin{aligned} X^f_{i,n+1} &= X^f_{i,n+\frac12} + f_i(X^f_{n+\frac12})\,\Delta t/2 + \sum_{j=1}^{D} g_{ij}(X^f_{n+\frac12})\,\delta w_{n+\frac12} \\ &\quad + \sum_{j,k=1}^{D} h_{ijk}(X^f_{n+\frac12})\left(\delta w_{j,n+\frac12}\,\delta w_{k,n+\frac12} - \Omega_{jk}\,\Delta t/2\right), \end{aligned} \tag{4.4}$$

in which

$$\delta w_n \equiv w(t_{n+\frac{1}{2}}) - w(t_n), \quad \delta w_{n+\frac{1}{2}} \equiv w(t_{n+1}) - w(t_{n+\frac{1}{2}}) \tag{4.5}$$

are the Brownian increments over the first and second halves of the coarse timestep, and so $\Delta w_n = \delta w_n + \delta w_{n+\frac{1}{2}}$.

The antithetic approximation X_n^a is defined by exactly the same discretisation except that the Brownian increments δw_n and $\delta w_{n+\frac{1}{2}}$ are swapped, so that

$$\begin{aligned}
X^a_{i,n+\frac{1}{2}} &= X^a_{i,n} + f_i(X^a_n)\,\Delta t/2 + \sum_{j=1}^{D} g_{ij}(X^a_n)\,\delta w_{n+\frac{1}{2}} \\
&\quad + \sum_{j,k=1}^{D} h_{ijk}(X^a_n)\left(\delta w_{j,n+\frac{1}{2}}\,\delta w_{k,n+\frac{1}{2}} - \Omega_{jk}\,\Delta t/2\right), \\
X^a_{i,n+1} &= X^a_{i,n+\frac{1}{2}} + f_i(X^a_{n+\frac{1}{2}})\,\Delta t/2 + \sum_{j=1}^{D} g_{ij}(X^a_{n+\frac{1}{2}})\,\delta w_{j,n} \\
&\quad + \sum_{j,k=1}^{D} h_{ijk}(X^a_{n+\frac{1}{2}})\left(\delta w_{j,n}\,\delta w_{k,n} - \Omega_{jk}\,\Delta t/2\right).
\end{aligned} \tag{4.6}$$

Since δw_n and $\delta w_{n+\frac{1}{2}}$ are independent and identically distributed, X^a has exactly the same distribution as X^f, and hence $\mathbb{E}[P(X^a)] = \mathbb{E}[P(X^f)]$. In addition, the following lemma follows directly from Lemma 4.2 and Lemma 4.4.

Lemma 4.5. *Let X^f and X^a be as defined above. Then for $p \geq 2$, there exists a constant K_p, independent of the time step, such that*

$$\begin{aligned}
\mathbb{E}\left[\max_{0\leq n\leq N}\left\|X_n^f\right\|^p\right] \leq K_p, &\qquad \max_{0\leq n<N}\mathbb{E}\left[\left\|X^f_{n+\frac{1}{2}} - X^f_n\right\|^p\right] \leq K_p\,\Delta t^{p/2}, \\
\mathbb{E}\left[\max_{0\leq n\leq N}\left\|X_n^a\right\|^p\right] \leq K_p, &\qquad \max_{0\leq n<N}\mathbb{E}\left[\left\|X^a_{n+\frac{1}{2}} - X^a_n\right\|^p\right] \leq K_p\,\Delta t^{p/2}.
\end{aligned}$$

4.3 Numerical analysis

The analysis is presented as a sequence of lemmas and theorems, with the proofs deferred to the Appendix. The outline is as follows:

- Lemma 4.6 bounds $\|X_n^f - X_n^a\|$ over a coarse timestep;
- Lemma 4.7 gives a representation of the discrete equation for X_n^f over a coarse timestep, and Corollary 4.8 gives the corresponding representation for X_n^a;
- Lemma 4.9 gives a representation of the discrete equation describing the evolution of the average $\overline{X}_n^f = \frac{1}{2}(X_n^f + X_n^a)$ over a coarse timestep;
- Theorem 4.10 bounds $\|\overline{X}_n^f - X_n^c\|$ over a coarse timestep.

Lemma 4.6. *For all integers $p \geq 2$, there exists a constant K_p such that*

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|X^f_n - X^a_n\|^p\right] \leq K_p\, \Delta t^{p/2}$$

Lemma 4.7. *Difference equation (4.4) for X^f_n can be expressed as*

$$\begin{aligned} X^f_{i,n+1} \;&=\; X^f_{i,n} + f_i(X^f_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(X^f_n)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(X^f_n)\,(\Delta w_{j,n}\Delta w_{k,n} - \Omega_{jk}\Delta t) \\ &\quad - \sum_{j,k=1}^{D} h_{ijk}(X^f_n)\left(\delta w_{j,n}\,\delta w_{k,n+\frac{1}{2}} - \delta w_{k,n}\,\delta w_{j,n+\frac{1}{2}}\right) \\ &\quad + M^f_{i,n} + N^f_{i,n}, \end{aligned}$$

where $\mathbb{E}[M^f_n \,|\, \mathcal{F}_n] = 0$, and for any integer $p\geq 2$ there exists a constant K_p such that

$$\max_{0\leq n\leq N} \mathbb{E}\left[\|M^f_n\|^p\right] \leq K_p\,\Delta t^{3p/2}, \qquad \max_{0\leq n\leq N} \mathbb{E}\left[\|N^f_n\|^p\right] \leq K_p\,\Delta t^{2p}.$$

Corollary 4.8. *Difference equation (4.6) for X^a_n can be expressed as*

$$\begin{aligned} X^a_{i,n+1} \;&=\; X^a_{i,n} + f_i(X^a_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(X^a_n)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(X^a_n)\,(\Delta w_{j,n}\Delta w_{k,n} - \Omega_{jk}\Delta t) \\ &\quad + \sum_{j,k=1}^{D} h_{ijk}(X^a_n)\left(\delta w_{j,n}\,\delta w_{k,n+\frac{1}{2}} - \delta w_{k,n}\,\delta w_{j,n+\frac{1}{2}}\right) \\ &\quad + M^a_{i,n} + N^a_{i,n}, \end{aligned}$$

where $\mathbb{E}[M^a_n \,|\, \mathcal{F}_n] = 0$, and for any integer $p\geq 2$ there exists a constant K_p such that

$$\max_{0\leq n\leq N} \mathbb{E}\left[\|M^a_n\|^p\right] \leq K_p\,\Delta t^{3p/2}, \qquad \max_{0\leq n\leq N} \mathbb{E}\left[\|N^a_n\|^p\right] \leq K_p\,\Delta t^{2p}.$$

Lemma 4.9. *The difference equation for $\overline{X}^f_n \equiv \frac{1}{2}(X^f_n + X^a_n)$ can be expressed as*

$$\begin{aligned} \overline{X}^f_{i,n+1} \;&=\; \overline{X}^f_{i,n} + f_i(\overline{X}^f_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(\overline{X}^f_n)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(\overline{X}^f_n)\,(\Delta w_{j,n}\,\Delta w_{k,n} - \Omega_{jk}\Delta t) \\ &\quad + M_{i,n} + N_{i,n}, \end{aligned}$$

where $\mathbb{E}[M_n \,|\, \mathcal{F}_n] = 0$, and for any integer $p\geq 2$ there exists a constant K_p such that

$$\max_{0\leq n\leq N} \mathbb{E}\left[\|M_n\|^p\right] \leq K_p\,\Delta t^{3p/2}, \qquad \max_{0\leq n\leq N} \mathbb{E}\left[\|N_n\|^p\right] \leq K_p\,\Delta t^{2p}.$$

Theorem 4.10. *For all $p \geq 2$, there exists a constant K_p such that*

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|\overline{X}^f_n - X^c_n\|^p\right] \leq K_p\,\Delta t^{p}.$$

4.4 Piecewise linear interpolation analysis

The piecewise linear interpolant $X^c(t)$ for the coarse path is defined within the coarse timestep interval $[t_k, t_{k+1}]$ as

$$X^c(t) \equiv (1-\lambda)\, X^c_k + \lambda\, X^c_{k+1}, \quad \lambda \equiv \frac{t - t_k}{t_{k+1} - t_k}.$$

Likewise, the piecewise linear interpolants $X^f(t)$ and $X^a(t)$ are defined on the fine timestep $[t_k, t_{k+\frac{1}{2}}]$ as

$$X^f(t) \equiv (1-\lambda)\, X^f_k + \lambda\, X^f_{k+\frac{1}{2}}, \quad X^a(t) \equiv (1-\lambda)\, X^a_k + \lambda\, X^a_{k+\frac{1}{2}}, \quad \lambda \equiv \frac{t - t_k}{t_{k+\frac{1}{2}} - t_k},$$

and there is a corresponding definition for the fine timestep $[t_{k+\frac{1}{2}}, t_{k+1}]$.

The proofs of the next two lemmas are in the Appendix, and the Theorem then follows directly.

Lemma 4.11. *For all integers $p \geq 2$, there exists a constant K_p such that*

$$\max_{0\leq n<N} \mathbb{E}\left[\|X^f_{n+\frac{1}{2}} - X^a_{n+\frac{1}{2}}\|^p \right] \leq K_p\, \Delta t^{p/2}.$$

Lemma 4.12. *For all $p \geq 2$, there exists a constant K_p such that*

$$\max_{0\leq n<N} \mathbb{E}\left[\left\|\overline{X}^f_{n+\frac{1}{2}} - X^c(t_{n+\frac{1}{2}})\right\|^p \right] \leq K_p\, \Delta t^p,$$

where $X^c(t_{n+\frac{1}{2}}) = \frac{1}{2}(X^c_n + X^c_{n+1})$ is the midpoint value of the coarse path interpolant.

Theorem 4.13. *For all $p \geq 2$, there exists a constant K_p such that*

$$\sup_{0\leq t\leq T} \mathbb{E}\left[\|X^f(t) - X^a(t)\|^p \right] \leq K_p\, \Delta t^{p/2},$$

$$\sup_{0\leq t\leq T} \mathbb{E}\left[\left\|\overline{X}^f(t) - X^c(t)\right\|^p \right] \leq K_p\, \Delta t^p,$$

where $\overline{X}^f(t)$ is the average of the piecewise linear interpolants $X^f(t)$ and $X^a(t)$.

5 European and Asian payoffs

5.1 European options

In the case of payoff which is a smooth function of the final state $x(T)$, taking $p\!=\!2$ in Lemma 2.2, $p = 4$ in Lemma 4.6 and $p = 2$ in Theorem 4.10, immediately gives the result that the multilevel variance

$$\mathbb{V}\left[\tfrac{1}{2}\left(P(X^f_N) + P(X^a_N) \right) - P(X^c_N) \right]$$

has an $O(\Delta t^2)$ upper bound. This matches the convergence rate for the multilevel method for scalar SDEs using the standard first order Milstein discretisation, and is much better than the $O(\Delta t)$ convergence obtained with the Euler-Maruyama discretisation.

However, very few financial payoff functions are twice differentiable on the entire domain $\mathbb{R}^d$. A more typical 2D example is a call option based on the minimum of two assets,

$$P(x(T)) \equiv \max\left(0, \min(x_1(T), x_2(T)) - K\right),$$

which is piecewise linear, with a discontinuity in the gradient along the three lines (s, K), (K, s) and (s, s) for $s \geq K$.

To handle such payoffs, we introduce a new assumption which bounds the probability of the solution of the SDE having a value at time T close to such lines with discontinuous gradients, and then formulate a theorem to show that the multilevel variance which results from using the antithetic estimator has an upper bound which is almost $O(\Delta t^{3/2})$.

Assumption 5.1. *The payoff function $P \in C(\mathbb{R}^d, \mathbb{R})$ has a uniform Lipschitz bound, so that there exists a constant L such that*

$$|P(x) - P(y)| \leq L\, |x - y|, \quad \forall\, x, y \in \mathbb{R}^d,$$

and the first and second derivatives exist, are continuous and have uniform bound L at all points $x \notin K$, where K is a set of zero measure, and there exists a constant c such that the probability of the SDE solution $x(T)$ being within a neighbourhood of the set K has the bound

$$\mathbb{P}\left(\min_{y \in K} \|x(T) - y\| \leq \varepsilon\right) \leq c\, \varepsilon, \quad \forall\, \varepsilon > 0.$$

In a 1D context, Assumption 5.1 corresponds to an assumption of a locally bounded density for $x(T)$.

Theorem 5.2. *If the SDE satisfies the conditions of Assumption 4.1, and the payoff satisfies Assumption 5.1, then*

$$\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(X_N^c)\right)^2\right] = o(\Delta t^{3/2-\delta})$$

for any $\delta > 0$.

Proof. We start by noting that

$$\begin{aligned}\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(X_N^c)\right)^2\right] \;\leq\; & 2\,\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f)\right)^2\right] \\ & + 2\,\mathbb{E}\left[\left(\tfrac{1}{2}(P(\overline{X}_N^f) - P(X_N^c)\right)^2\right].\end{aligned}$$

The second term on the right-hand-side has an $O(\Delta t^2)$ bound due to the uniform Lipschitz bound for the payoff, together with the result from Theorem 4.10 for $p = 2$.

The objective now is to prove that the first term has a $o(\Delta t^{3/2-\delta})$ bound for any $\delta > 0$. The analysis follows the approach used in [7]. To prove this for a particular value of δ, we define

$\varepsilon = \Delta t^{1/2-\delta/2}$, and consider the three events

$$\begin{aligned} A &\equiv \left\{ \min_{y\in K} \|x(T) - y\| \leq \varepsilon \right\}, \\ B &\equiv \left\{ \|x(T) - X_N^f\| \geq \tfrac{1}{2}\varepsilon \right\}, \\ C &\equiv \left\{ \|X_N^f - X_N^a\| \geq \tfrac{1}{2}\varepsilon \right\}. \end{aligned}$$

Using $\mathbf{1}_A$ to indicate the indicator function for event A, and A^c to denote the complement of A, we have

$$\begin{aligned} \mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) \right)^2 \right] &= \mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) \right)^2 \mathbf{1}_{A\cup B\cup C} \right] \\ &+ \mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) \right)^2 \mathbf{1}_{A^c\cap B^c\cap C^c} \right]. \end{aligned}$$

Looking at the first of the two terms on the right-hand-side, then Hölder's inequality gives

$$\begin{aligned} &\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) \right)^2 \mathbf{1}_{A\cup B\cup C} \right] \\ &\leq \mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) \right)^{2p} \right]^{1/p} (\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C))^{1/q} \end{aligned}$$

for any $p, q \geq 1$, with $p^{-1} + q^{-1} = 1$. The Markov inequality gives,

$$\mathbb{P}(B) \leq \mathbb{E}\left[\left\| x(T) - X_N^f \right\|^m \right] \ / \ (\tfrac{1}{2}\varepsilon)^m,$$

for any $m \geq 1$. Using the strong convergence property from Lemma 4.2, and the definition of ε, we can take m to be sufficiently large so that

$$\tfrac{1}{2}m - \tfrac{1-\delta}{2}\, m > \tfrac{1-\delta}{2}$$

and hence there exists a constant c_1 such that $\mathbb{P}(B) \leq c_1\, \varepsilon$. Using Lemma 4.6, one can obtain a similar bound $\mathbb{P}(C) \leq c_2\, \varepsilon$, and then q can be chosen sufficiently close to 1 so that

$$(\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C))^{1/q} \ \leq \ (1 + c_1 + c_2)^{1/q} \Delta t^{(1/2-\delta/2)/q} \ = \ o(\Delta t^{1/2-\delta}).$$

Since

$$\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) = \tfrac{1}{2}(P(X_N^f) - P(\overline{X}_N^f)) + \tfrac{1}{2}(P(X_N^a) - P(\overline{X}_N^f)),$$

the uniform Lipschitz bound gives

$$\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) \right)^{2p} \right]^{1/p} \leq \ L^2\ \mathbb{E}\left[\left\| X_N^f - X_N^a \right\|^{2p} \right]^{1/p} \leq \ c_3\, \Delta t,$$

for some constant c_3 due to Lemma 4.6, and hence

$$\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f) + P(X_N^a)) - P(\overline{X}_N^f) \right)^2 \mathbf{1}_{A\cup B\cup C} \right] = o(\Delta t^{3/2-\delta}).$$

Lastly, we consider the second term

$$\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f)+P(X_N^a))-P(\overline{X}_N^f)\right)^2 \mathbf{1}_{A^c\cap B^c\cap C^c}\right].$$

Given a path sample $\omega\in(B^c\cap C^c)$, if the straight line between X_N^f and X_N^a contains a point $y\in K$, then $\|y-X_N^f\|$ and $\|x(T)-X_N^f\|$ are both less than $\varepsilon/2$, and hence $\|x(T)-y\|<\varepsilon$.

Thus, for a path sample $\omega\in(A^c\cap B^c\cap C^c)$, the straight line between X_N^f and X_N^a does not contain any points in K. It is therefore possible to perform a second order truncated Taylor expansion as in the proof of Lemma 2.2, and deduce that there exists a constant c_4 such that

$$\mathbb{E}\left[\left(\tfrac{1}{2}(P(X_N^f)+P(X_N^a))-P(\overline{X}_N^f)\right)^2 \mathbf{1}_{A^c\cap B^c\cap C^c}\right] \leq c_4\, \mathbb{E}\left[\left\|X_N^f-X_N^a\right\|^4\right],$$

which has an $O(\Delta t^2)$ bound due to Lemma 4.6. □

5.2 Asian payoffs

For an Asian option, the payoff depends on the average

$$x_{ave} \;\equiv\; T^{-1}\int_0^T x(t)\,\mathrm{d}t.$$

This can be approximated by integrating the appropriate piecewise linear interpolant which gives

$$\begin{aligned}
X_{ave}^c &\equiv T^{-1}\int_0^T X^c(t)\,\mathrm{d}t &&= N^{-1}\sum_{n=0}^{N-1}\tfrac{1}{2}(X_n^c+X_{n+1}^c),\\
X_{ave}^f &\equiv T^{-1}\int_0^T X^f(t)\,\mathrm{d}t &&= N^{-1}\sum_{n=0}^{N-1}\tfrac{1}{4}(X_n^f+2X_{n+\frac{1}{2}}^f+X_{n+1}^f),\\
X_{ave}^a &\equiv T^{-1}\int_0^T X^a(t)\,\mathrm{d}t &&= N^{-1}\sum_{n=0}^{N-1}\tfrac{1}{4}(X_n^a+2X_{n+\frac{1}{2}}^a+X_{n+1}^a).
\end{aligned}$$

Due to Hölder's inequality,

$$\mathbb{E}\left[\left\|X_{ave}^f-X_{ave}^a\right\|^p\right] \;\leq\; T^{-1}\int_0^T \mathbb{E}\left[\left\|X^f(t)-X^a(t)\right\|^p\right]\,\mathrm{d}t \;\leq\; \sup_{[0,T]}\mathbb{E}\left[\left\|X^f(t)-X^a(t)\right\|^p\right],$$

and similarly

$$\mathbb{E}\left[\left\|\tfrac{1}{2}(X_{ave}^f+X_{ave}^a)-X_{ave}^c\right\|^p\right] \;\leq\; \sup_{[0,T]}\mathbb{E}\left[\left\|\overline{X}^f(t)-X^c(t)\right\|^p\right],$$

Hence, if the Asian payoff is a smooth function of the average, then taking $p=2$ in Lemma 2.2, $p=4$ in Corollary 4.11 and $p=2$ in Corollary 4.12, again gives a second order bound for the multilevel correction variance.

This analysis can be extended to include payoffs which are a smooth function of a number of intermediate variables, each of which is a linear functional of the path $x(t)$ of the form

$$\int_0^T g^T(t)\; x(t)\; \mu(\mathrm{d}t),$$

for some vector function $g(t)$ and measure $\mu(\mathrm{d}t)$. This includes weighted averages of $x(t)$ at a number of discrete times, as well as continuously-weighted averages over the whole time interval.

As with the European options, the analysis can also be extended to payoffs which are Lipschitz functions of the average, and have first and second derivatives which exist, and are continuous and uniformly bounded, except for a set of points K of zero measure.

Assumption 5.3. *The payoff $P \in C(\mathbb{R}^d, \mathbb{R})$ has a uniform Lipschitz bound, so that there exists a constant L such that*

$$|P(x) - P(y)| \leq L\; |x - y|\,, \quad \forall\, x, y \in \mathbb{R}^d,$$

and the first and second derivatives exist, are continuous and have uniform bound L at all points $x \notin K$, where K is a set of zero measure, and there exists a constant c such that the probability of x_{ave} being within a neighbourhood of the set K has the bound

$$\mathbb{P}\left(\min_{y\in K} \|x_{ave} - y\| \leq \varepsilon\right) \leq c\;\varepsilon, \quad \forall\, \varepsilon > 0.$$

Theorem 5.4. *If the SDE satisfies the conditions of Assumption 4.1, and the payoff satisfies Assumption 5.3, then*

$$\mathbb{E}\left[\left(\tfrac{1}{2}(P(X^f_{ave}) + P(X^a_{ave})) - P(X^c_{ave})\right)^2\right] = o(\Delta t^{3/2-\delta})$$

for any $\delta > 0$.

6 Conclusions

In this paper we have constructed a new antithetic multilevel Monte Carlo estimator for multi-dimensional SDEs, with a variance which is $O(\Delta t^2)$ when the payoff function is smooth, and almost an $O(\Delta t^{3/2})$ when it is Lipschitz and piecewise smooth. The algorithm is very easy to implement; all that is required is to calculate a second fine path for which the odd and even Brownian increments are swapped.

In the European and Asian payoff cases considered in this paper, it reduces the computational complexity for an ϵ root-mean-square error to $O(\epsilon^{-2})$, compared to $O(\epsilon^{-2}(\log \epsilon)^2)$ for the multilevel method using the Euler-Maruyama discretisation, and $O(\epsilon^{-3})$ for the standard Monte Carlo method. Furthermore, by ensuring that the dominant computational effort is on the coarsest levels (since $\beta > 1$), it is now feasible to obtain further improvements using quasi-Monte Carlo techniques [9].

In a second paper, we will extend the analysis to cover digital and barrier options. The improvements from an extended version of the antithetic treatment are then more substantial, improving the complexity from $O(\epsilon^{-5/2})$ to approximately $O(\epsilon^{-2})$.

A Proof of main results

A.1 Proof of Lemma 4.6

Proof. Conditional on the Brownian increments Δw for the coarse path X^c, the Brownian increments for X^f and X^a have exactly the same distribution, and therefore $X^a_n - X^c_n$ has exactly the same distribution as $X^f_n - X^c_n$. Hence we obtain, using inequality (2.4),

$$\begin{aligned}
\mathbb{E}\left[\max_{0\le n\le N} \|X^f_n - X^a_n\|^p\right] &\le 2^{p-1}\left(\mathbb{E}\left[\max_{0\le n\le N} \|X^f_n - X^c_n\|^p\right] + \mathbb{E}\left[\max_{0\le n\le N} \|X^a_n - X^c_n\|^p\right]\right)\\
&= 2^p\, \mathbb{E}\left[\max_{0\le n\le N} \|X^f_n - X^c_n\|^p\right]\\
&\le 2^{2p-1}\left(\mathbb{E}\left[\max_{0\le n\le N} \|X^f_n - x(t_n)\|^p\right] + \mathbb{E}\left[\max_{0\le n\le N} \|X^c_n - x(t_n)\|^p\right]\right).
\end{aligned}$$

The desired result then follows from the strong convergence property in Lemma 4.2. $\square$

A.2 Proof of Lemma 4.7 and Corollary 4.8

Proof. Combining the two equations in (4.3), and using the identity

$$\Delta w_{j,n}\, \Delta w_{k,n} = (\delta w_{j,n} + \delta w_{j,n+\frac{1}{2}})\, (\delta w_{k,n} + \delta w_{k,n+\frac{1}{2}})$$

together with the definition of h_{ijk} in (1.2) gives, after considerable re-arrangement,

$$\begin{aligned}
X^f_{i,n+1} &= X^f_{i,n} + f_i(X^f_n)\,\Delta t + \sum_{j=1}^{D} g_{ij}(X^f_n)\,\Delta w_{j,n}\\
&\quad + \sum_{j,k=1}^{D} h_{ijk}(X^f_n)\,(\Delta w_{j,n}\,\Delta w_{k,n} - \Omega_{jk}\Delta t)\\
&\quad - \sum_{j,k=1}^{D} h_{ijk}(X^f_n)\left(\delta w_{j,n}\,\delta w_{k,n+\frac{1}{2}} - \delta w_{k,n}\,\delta w_{j,n+\frac{1}{2}}\right)\\
&\quad + R_{i,n} + M^{(2)}_{i,n} + M^{(3)}_{i,n},
\end{aligned}$$

where

$$\begin{aligned}
R_{i,n} &= \left(f_i(X^f_{n+\frac{1}{2}}) - f_i(X^f_n)\right)\Delta t/2,\\
M^{(2)}_{i,n} &= \sum_{j=1}^{D}\left(g_{ij}(X^f_{n+\frac{1}{2}}) - g_{ij}(X^f_n) - 2\sum_{k=1}^{D} h_{ijk}(X^f_n)\delta w_{k,n}\right)\delta w_{j,n+\frac{1}{2}},\\
M^{(3)}_{i,n} &= \sum_{j,k=1}^{D}\left(h_{ijk}(X^f_{n+\frac{1}{2}}) - h_{ijk}(X^f_n)\right)\left(\delta w_{j,n+\frac{1}{2}}\delta w_{k,n+\frac{1}{2}} - \Omega_{jk}\Delta t/2\right).
\end{aligned}$$

Considering R_n, a Taylor expansion gives

$$\begin{aligned} f_i(X^f_{n+\frac{1}{2}}) - f_i(X^f_n) &= \sum_{j=1}^{d} \frac{\partial f_i}{\partial x_j}(X^f_n)\ \left(X^f_{j,n+\frac{1}{2}} - X^f_{j,n}\right) \\ &\quad + \frac{1}{2}\sum_{j,k=1}^{d} \frac{\partial^2 f_i}{\partial x_j \partial x_k}(\xi_1)\ \left(X^f_{j,n+\frac{1}{2}} - X^f_{j,n}\right)\left(X^f_{k,n+\frac{1}{2}} - X^f_{k,n}\right) \end{aligned}$$

for some ξ_1 which lies on the line between X^f_n and $X^f_{n+\frac{1}{2}}$. Hence, R_n can be split into two parts, $R_n = M^{(1)}_n + N_n$, where

$$M^{(1)}_{i,n} = \sum_{j=1}^{d}\sum_{k=1}^{D} \frac{\partial f_i}{\partial x_j}(X^f_n)\ g_{jk}(X^f_n)\,\delta w_{k,n}\ \Delta t/2,$$

and

$$\begin{aligned} N_{i,n} &= \sum_{j=1}^{d} \frac{\partial f_i}{\partial x_j}(X^f_n)\left(f_j(X^f_n)\ \Delta t/2 + \sum_{k,l=1}^{D} h_{jkl}(X^f_n)\,(\delta w_{k,n}\delta w_{l,n} - \Omega_{kl}\Delta t/2)\right)\Delta t/2 \\ &\quad + \frac{1}{2}\sum_{j,k=1}^{d} \frac{\partial^2 f_i}{\partial x_j \partial x_k}(\xi_1)\ \left(X^f_{j,n+\frac{1}{2}} - X^f_{j,n}\right)\left(X^f_{k,n+\frac{1}{2}} - X^f_{k,n}\right)\ \Delta t/2. \end{aligned}$$

Considering $M^{(2)}_n$, a Taylor expansion gives

$$\begin{aligned} g_{ij}(X^f_{n+\frac{1}{2}}) - g_{ij}(X^f_n) &= \sum_{k=1}^{d} \frac{\partial g_{ij}}{\partial x_k}(X^f_n)\ \left(X^f_{k,n+\frac{1}{2}} - X^f_{k,n}\right) \\ &\quad + \frac{1}{2}\sum_{k,l=1}^{d} \frac{\partial^2 g_{ij}}{\partial x_k \partial x_l}(\xi_2)\ \left(X^f_{k,n+\frac{1}{2}} - X^f_{k,n}\right)\left(X^f_{l,n+\frac{1}{2}} - X^f_{l,n}\right), \end{aligned}$$

for some ξ_2 on the line between X^f_n and $X^f_{n+\frac{1}{2}}$, and therefore

$$\begin{aligned} M^{(2)}_{i,n} &= \sum_{j=1}^{D}\sum_{k=1}^{d} \frac{\partial g_{ij}}{\partial x_k}(X^f_n)\left(f_k(X^f_n)\,\Delta t/2 + \sum_{l,m=1}^{D} h_{klm}(X^f_n)\,(\delta w_{l,n}\delta w_{m,n} - \Omega_{lm}\Delta t/2)\right)\delta w_{j,n+\frac{1}{2}} \\ &\quad + \frac{1}{2}\sum_{j=1}^{D}\sum_{k,l=1}^{d} \frac{\partial^2 g_{ij}}{\partial x_k \partial x_l}(\xi_2)\ \left(X^f_{k,n+\frac{1}{2}} - X^f_{k,n}\right)\left(X^f_{l,n+\frac{1}{2}} - X^f_{l,n}\right)\delta w_{j,n+\frac{1}{2}}. \end{aligned}$$

Finally, considering $M^{(3)}_n$ we have

$$\begin{aligned} M^{(3)}_{i,n} &= \sum_{j,k=1}^{D} \left(h_{ijk}(X^f_{n+\frac{1}{2}}) - h_{ijk}(X^f_n)\right)\left(\delta w_{j,n+\frac{1}{2}}\delta w_{k,n+\frac{1}{2}} - \Omega_{jk}\Delta t/2\right) \\ &= \sum_{j,k=1}^{D}\sum_{l=1}^{d} \frac{\partial h_{ijk}}{\partial x_l}(\xi_3)\left(X^f_{l,n+\frac{1}{2}} - X^f_{l,n}\right)\left(\delta w_{j,n+\frac{1}{2}}\delta w_{k,n+\frac{1}{2}} - \Omega_{jk}\Delta t/2\right), \end{aligned}$$

for some ξ_3 on the line between X^f_n and $X^f_{n+\frac{1}{2}}$.

Setting $M_n^f \equiv M_n^{(1)} + M_n^{(2)} + M_n^{(3)}$, it is clear that $\mathbb{E}[M_n^f \,|\, \mathcal{F}_n] = 0$ since δw_n is independent of X_n^f, and $\delta w_{n+\frac{1}{2}}$ is independent of X_n^f and $X_{n+\frac{1}{2}}^f$.

All that remains is to bound the magnitude of $\mathbb{E}[\,\|M_n^f\|^p]$ and $\mathbb{E}[\,\|N_n^f\|^p]$. Looking at two of the terms in $M_{i,n}^{(2)}$, for example, the uniform bound on the first derivatives of g, together with the fact that $\delta w_{n+\frac{1}{2}}$ is independent of both X_n^f and δw_n leads to

$$
\begin{aligned}
\mathbb{E}&\left[\left|\frac{\partial g_{ij}}{\partial x_k}(X_n^f)\; h_{klm}(X_n^f)\; \delta w_{l,n}\, \delta w_{m,n}\, \delta w_{j,n+\frac{1}{2}}\right|^p\right] \\
&\qquad\qquad \leq\;\; L^p\; \mathbb{E}\left[\left|h_{klm}(X_n^f)\right|^p\right]\; \mathbb{E}\left[\,\|\delta w_n\|^{2p}\right] \mathbb{E}\left[\left\|\delta w_{n+\frac{1}{2}}\right\|^p\right],
\end{aligned}
$$

and the uniform bound on the second derivatives of g, together with the fact that $\delta w_{n+\frac{1}{2}}$ is independent of both X_n^f and $X_{n+\frac{1}{2}}^f$ leads to

$$
\begin{aligned}
\mathbb{E}&\left[\left|\frac{\partial^2 g_{ij}}{\partial x_k \partial x_l}(\xi_2)\; \left(X_{k,n+\frac{1}{2}}^f - X_{k,n}^f\right) \left(X_{l,n+\frac{1}{2}}^f - X_{l,n}^f\right) \delta w_{j,n+\frac{1}{2}}.\right|^p\right] \\
&\qquad\qquad \leq\;\; L^p\; \mathbb{E}\left[\left\|X_{n+\frac{1}{2}}^f - X_n^f\right\|^{2p}\right]\; \mathbb{E}\left[\left\|\delta w_{n+\frac{1}{2}}\right\|^p\right].
\end{aligned}
$$

Combining the uniform bound on $\mathbb{E}\left[\left|h_{ijk}(X_n^f)\right|^{2p}\right]$ from Corollary 4.3 with the bounds from Lemma 4.4, and standard results for the moments of Brownian increments, gives the required $O(\Delta t^{3p/2})$ bound for each of the two terms considered.

Deriving similar bounds for the other terms in M^f and N^f, and combining them using (2.4), eventually gives the desired bounds for both $\mathbb{E}[\,\|M_n^f\|^p]$ and $\mathbb{E}[\,\|N_n^f\|^p]$.

The proof is almost exactly the same for Corollary 4.8. The sign change in the second line of the equation in the statement of the Corollary is due to the swapping of the Brownian increments for the first and second halves of the timestep.

□

A.3 Proof of Lemma 4.9

Proof. Recalling that $\overline{X}^f = \frac{1}{2}(X^f + X^a)$, taking the average of the results from Lemma 4.7 and Corollary 4.8 gives

$$
\begin{aligned}
\overline{X}_{i,n+1}^f \;\;=\;\; & \overline{X}_{i,n}^f + f_i(\overline{X}_n^f)\,\Delta t + \sum_{j=1}^{D} g_{ij}(\overline{X}_n^f)\,\Delta w_{j,n} + \sum_{j,k=1}^{D} h_{ijk}(\overline{X}_n^f)\,(\Delta w_{j,n}\Delta w_{k,n} - \Omega_{jk}\Delta t) \\
& + \tfrac{1}{2}\left(M_{i,n}^f + N_{i,n}^f + M_{i,n}^a + N_{i,n}^a\right) \;+\; M_{i,n}^{(1)} + M_{i,n}^{(2)} + M_{i,n}^{(3)} + N_{i,n}^{(1)},
\end{aligned}
$$

where

$$
\begin{aligned}
N^{(1)}_{i,n} &= \left(\tfrac{1}{2}(f_i(X^f_n)+f_i(X^a_n)) - f_i(\overline{X}^f_n)\right)\,\Delta t,\\
M^{(1)}_{i,n} &= \sum_{j=1}^{D} \left(\tfrac{1}{2}(g_{ij}(X^f_n)+g_{ij}(X^a_n)) - g_{ij}(\overline{X}^f_n)\right)\,\Delta w_{j,n},\\
M^{(2)}_{i,n} &= \sum_{j,k=1}^{D} \left(\tfrac{1}{2}(h_{ijk}(X^f_n)+h_{ijk}(X^a_n)) - h_{ijk}(\overline{X}^f_n)\right)\,(\Delta w_{j,n}\Delta w_{k,n} - \Omega_{jk}\Delta t)\,,\\
M^{(3)}_{i,n} &= \sum_{j,k=1}^{D} \tfrac{1}{2}\left(h_{ijk}(X^f_n)-h_{ijk}(X^a_n)\right)\,\left(\delta w_{j,n}\,\delta w_{k,n+\frac{1}{2}} - \delta w_{k,n}\,\delta w_{j,n+\frac{1}{2}}\right).
\end{aligned}
$$

Setting

$$
\begin{aligned}
M_n &= \tfrac{1}{2}\left(M^f_n + M^a_n\right) + M^{(1)}_n + M^{(2)}_n + M^{(3)}_n,\\
N_n &= \tfrac{1}{2}\left(N^f_n + N^a_n\right) + N^{(1)}_n,
\end{aligned}
$$

it is clear that $\mathbb{E}[M_n\,|\,\mathcal{F}_n] = 0$, and all that remains is to bound the magnitude of $\mathbb{E}[\,\|M_n\|^p]$ and $\mathbb{E}[\,\|N_n\|^p]$. By performing second order Taylor series expansions for $f(x)$ and $g(x)$, and first order expansions for $h(x)$, all about $\overline{X}^f_n$, we obtain

$$
\begin{aligned}
N^{(1)}_{i,n} &= \frac{1}{16}\sum_{j,k=1}^{d}\left(\frac{\partial^2 f_i}{\partial x_j \partial x_k}(\xi_1) + \frac{\partial^2 f_i}{\partial x_j \partial x_k}(\xi_2)\right)(X^f_{j,n} - X^a_{j,n})\,(X^f_{k,n} - X^a_{k,n})\,\Delta t,\\
M^{(1)}_{i,n} &= \frac{1}{16}\sum_{j=1}^{D}\sum_{k,l=1}^{d}\left(\frac{\partial^2 g_{ij}}{\partial x_k \partial x_l}(\xi_3) + \frac{\partial^2 g_{ij}}{\partial x_k \partial x_l}(\xi_4)\right)(X^f_{k,n} - X^a_{k,n})\,(X^f_{l,n} - X^a_{l,n})\,\Delta w_{j,n},\\
M^{(2)}_{i,n} &= \frac{1}{4}\sum_{j,k=1}^{D}\sum_{l=1}^{d}\left(\frac{\partial h_{ijk}}{\partial x_l}(\xi_5) - \frac{\partial h_{ijk}}{\partial x_l}(\xi_6)\right)(X^f_{l,n} - X^a_{l,n})\,(\Delta w_{j,n}\Delta w_{k,n} - \Omega_{jk}\Delta t)\,,\\
M^{(3)}_{i,n} &= \frac{1}{4}\sum_{j,k=1}^{D}\sum_{l=1}^{d}\left(\frac{\partial h_{ijk}}{\partial x_l}(\xi_7) + \frac{\partial h_{ijk}}{\partial x_l}(\xi_8)\right)(X^f_{l,n} - X^a_{l,n})\left(\delta w_{j,n}\,\delta w_{k,n+\frac{1}{2}} - \delta w_{k,n}\,\delta w_{j,n+\frac{1}{2}}\right),
\end{aligned}
$$

for some $\xi_1, \xi_3, \xi_5, \xi_7$ between $\overline{X}^f_n$ and X^f_n, and $\xi_2, \xi_4, \xi_6, \xi_8$ between $\overline{X}^f_n$ and X^a_n.

Using the same arguments as in the final part of the proof of Lemma 4.7, together with the bounds on $\mathbb{E}[\,\|M^f_n\|^p]$, $\mathbb{E}[\,\|M^a_n\|^p]$, $\mathbb{E}[\,\|N^f_n\|^p]$ and $\mathbb{E}[\,\|N^a_n\|^p]$, leads to the required bounds for the moments of M_n and N_n. □

A.4 Proof of Theorem 4.10

Proof. If we define $S_n = \mathbb{E}\left[\max_{m\leq n} \|\overline{X}^f_m - X^c_m\|^p\right]$ then inequality (2.4) gives

$$
S_n \;\leq\; d^{p-1}\sum_{i=1}^{d}\mathbb{E}\left[\max_{m\leq n}\left|\overline{X}^f_{i,m} - X^c_{i,m}\right|^p\right]. \tag{A.1}
$$

Taking the difference between the equation in Lemma 4.9 and equation (4.2), and summing over the first m timesteps we obtain

$$
\begin{aligned}
\overline{X}^f_{i,m} - X^c_{i,m} \;=\; & \sum_{l=0}^{m-1} \left(f_i(\overline{X}^f_{i,l}) - f_i(X^c_{i,l}) \right) \Delta t \\
+ & \sum_{l=0}^{m-1} \sum_{j=1}^{D} \left(g_{ij}(\overline{X}^f_{i,l}) - g_{ij}(X^c_{i,l}) \right) \Delta w_{j,l} \\
+ & \sum_{l=0}^{m-1} \sum_{j,k=1}^{D} \left(h_{ijk}(\overline{X}^f_{i,l}) - h_{ijk}(X^c_{i,l}) \right) \left(\Delta w_{j,l} \Delta w_{k,l} - \Omega_{jk} \Delta t \right) \\
+ & \sum_{l=0}^{m-1} M_{i,l} \;+\; \sum_{l=0}^{m-1} N_{i,l},
\end{aligned}
$$

and using inequality (2.4) again gives

$$
\begin{aligned}
& \mathbb{E}\left[\max_{m \le n} \left| \overline{X}^f_{i,m} - X^c_{i,m} \right|^p \right] \\
& \quad \le \; 5^{p-1} \Bigg(\mathbb{E}\left[\max_{m \le n} \left| \sum_{l=0}^{m-1} \left(f_i(\overline{X}^f_{i,l}) - f_i(X^c_{i,l}) \right) \Delta t \right|^p \right] \\
& \qquad + \mathbb{E}\left[\max_{m \le n} \left| \sum_{l=0}^{m-1} \sum_{j=1}^{D} \left(g_{ij}(\overline{X}^f_{i,l}) - g_{ij}(X^c_{i,l}) \right) \Delta w_{j,l} \right|^p \right] \\
& \qquad + \mathbb{E}\left[\max_{m \le n} \left| \sum_{l=0}^{m-1} \sum_{j,k=1}^{D} \left(h_{ijk}(\overline{X}^f_{i,l}) - h_{ijk}(X^c_{i,l}) \right) \left(\Delta w_{j,l} \Delta w_{k,l} - \Omega_{jk} \Delta t \right) \right|^p \right] \\
& \qquad + \mathbb{E}\left[\max_{m \le n} \left| \sum_{l=0}^{m-1} M_{i,l} \right|^p \right] + \mathbb{E}\left[\max_{m \le n} \left| \sum_{l=0}^{m-1} N_{i,l} \right|^p \right] \Bigg). \qquad \text{(A.2)}
\end{aligned}
$$

We now need to bound each of the five expectations on the right-hand side of (A.2). The last is the easiest, since

$$
\left| \sum_{l=0}^{m-1} N_{i,l} \right|^p \;\le\; m^{p-1} \sum_{l=0}^{m-1} |N_{i,l}|^p \;\le\; n^{p-1} \sum_{l=0}^{n-1} |N_{i,l}|^p \,,
$$

and therefore

$$
\mathbb{E}\left[\max_{m \le n} \left| \sum_{l=0}^{m-1} N_{i,l} \right|^p \right] \;\le\; n^{p-1} \sum_{l=0}^{n-1} \mathbb{E}\left[\, |N_{i,l}|^p \right] \;\le\; c_1 (n \Delta t)^p \, \Delta t^p
$$

for some constant c_1 (which like other such constants in this proof will depend on p, L and T but not on Δt) due to Lemma 4.9.

Similarly, there exists a constant c_2 such that

$$
\begin{aligned}
\mathbb{E}\left[\max_{m \le n} \left| \sum_{l=0}^{m-1} \left(f_i(\overline{X}^f_{i,l}) - f_i(X^c_{i,l}) \right) \Delta t \right|^p \right] \;\le\; & n^{p-1} \sum_{l=0}^{m-1} \mathbb{E}\left[\left| f_i(\overline{X}^f_{i,l}) - f_i(X^c_{i,l}) \right|^p \right] \Delta t^p \\
\le\; & c_2 \, (n \, \Delta t)^{p-1} \sum_{m=0}^{n-1} S_m \, \Delta t,
\end{aligned}
$$

with the second step being due to the uniform bound on the first derivatives of f.

The other three expectations in (A.2) involve martingales, and so we can use the discrete Burkholder-Davis-Gundy inequality [1]. Starting again with the easiest, there are constants c_3, c_4 such that

$$\mathbb{E}\left[\max_{m\leq n}\left|\sum_{l=0}^{m-1} M_{i,l}\right|^p\right] \leq c_3\,\mathbb{E}\left[\left(\sum_{m=0}^{n-1}(M_{i,m})^2\right)^{p/2}\right] \leq c_3\, n^{p/2-1}\sum_{m=0}^{n-1}\mathbb{E}[\,|M_{i,m}|^p] \leq c_4\,(n\,\Delta t)^{p/2}\Delta t^p.$$

with the final step being due to Lemma 4.9.

Similarly, there exists a constant c_5 such that

$$\begin{aligned}
&\mathbb{E}\left[\max_{m\leq n}\left|\sum_{l=0}^{m-1}\sum_{j=1}^{D}\left(g_{ij}(\overline{X}^f_{i,l}) - g_{ij}(X^c_{i,l})\right)\Delta w_{j,l}\right|^p\right]\\
&\qquad\leq\quad c_5\, n^{p/2-1}\, D^{p-1}\sum_{m=0}^{n-1}\sum_{j=1}^{D}\mathbb{E}\left[\,\left|\left(g_{ij}(\overline{X}^f_{i,m}) - g_{ij}(X^c_{i,m})\right)\Delta w_{j,m}\right|^p\right].
\end{aligned}$$

Since $\Delta w_{j,m}$ is independent of both $\overline{X}^f_{i,m}$ and $X^c_{i,m}$, it follows that

$$\mathbb{E}\left[\,\left|\left(g_{ij}(\overline{X}^f_{i,m}) - g_{ij}(X^c_{i,m})\right)\Delta w_{j,m}\right|^p\right] = \mathbb{E}\left[\,\left|g_{ij}(\overline{X}^f_{i,m}) - g_{ij}(X^c_{i,m})\right|^p\right]\;\mathbb{E}\left[\,|\Delta w_{j,m}|^p\right].$$

Hence, because of the uniformly bounded first derivatives of g, and standard results for the moments of Brownian increments, there exists a constant c_6 such that

$$\mathbb{E}\left[\max_{m\leq n}\left|\sum_{l=0}^{m-1}\sum_{j=1}^{D}\left(g_{ij}(\overline{X}^f_{i,l}) - g_{ij}(X^c_{i,l})\right)\Delta w_{j,l}\right|^p\right] \leq c_6\,(n\,\Delta t)^{p/2-1}\sum_{m=0}^{n-1} S_m\,\Delta t.$$

Finally, following the same approach, there exists a constant c_7 such that

$$\begin{aligned}
&\mathbb{E}\left[\max_{m\leq n}\left|\sum_{l=0}^{m-1}\sum_{j,k=1}^{D}\left(h_{ijk}(\overline{X}^f_{i,l}) - h_{ijk}(X^c_{i,l})\right)\,(\Delta w_{j,l}\Delta w_{k,l} - \Omega_{jk}\Delta t)\right|^p\right]\\
&\qquad\qquad\leq\quad c_5\,(n\,\Delta t)^{p/2-1}\,\Delta t^{p/2}\sum_{m=0}^{n-1} S_m\,\Delta t.
\end{aligned}$$

Since $n\Delta t \leq T$ in all of the above inequalities, combining the above bounds for each term in (A.2), and inserting these into (A.1), there then exists a constant c_8 such that

$$S_n \leq c_8\left(\Delta t^p + \sum_{m=0}^{n-1} S_m\,\Delta t\right).$$

The desired result is then obtained from a discrete Grönwall inequality. $\square$

A.5 Proof of Lemma 4.11

Proof. The identity $X^f_{n+\frac12} - X^a_{n+\frac12} = (X^f_{n+\frac12} - X^f_n) + (X^f_n - X^a_n) + (X^a_n - X^a_{n+\frac12})$ gives

$$\left\| X^f_{n+\frac12} - X^a_{n+\frac12} \right\|^p \le 3^{p-1} \left(\left\| X^f_{n+\frac12} - X^f_n \right\|^p + \left\| X^f_n - X^a_n \right\|^p + \left\| X^a_{n+\frac12} - X^a_n \right\|^p \right).$$

It then follows from Lemma 4.5 and Lemma 4.6 that there exists a constant K_p, independent of both Δt and n, for which

$$\mathbb{E}\left[\left\| X^f_{n+\frac12} - X^a_{n+\frac12} \right\|^p \right] \le K_p \, \Delta t^{p/2}.$$

□

A.6 Proof of Lemma 4.12

Averaging the discrete equations for $X^f_{n+\frac12}$ and $X^a_{n+\frac12}$, and using the identities $\delta w_n = \frac12 \Delta w_n + \frac12 (\delta w_n - \delta w_{n+\frac12})$ and $\delta w_{n+\frac12} = \frac12 \Delta w_n - \frac12 (\delta w_n - \delta w_{n+\frac12})$, gives

$$\overline{X}^f_{i,n+\frac12} = \overline{X}^f_{i,n} + \tfrac12 \, f_i(\overline{X}^f_n) \, \Delta t + \tfrac12 \sum_{j=1}^{D} g_{ij}(\overline{X}^f_n) \, \Delta w_{j,n} + N_{i,n}, \tag{A.3}$$

where

$$\begin{aligned} N_{i,n} \;=\; & \tfrac12 \left(\tfrac12 (f_i(X^f_n) + f_i(X^a_n)) - f_i(\overline{X}^f_n) \right) \Delta t \\ & + \tfrac12 \sum_{j=1}^{D} \left(\tfrac12 (g_{ij}(X^f_n) + g_{ij}(X^a_n)) - g_{ij}(\overline{X}^f_n) \right) \Delta w_{j,n} \\ & + \tfrac14 \sum_{j=1}^{D} (g_{ij}(X^f_n) - g_{ij}(X^a_n)) \, (\delta w_{j,n} - \delta w_{j,n+\frac12}) \\ & + \tfrac12 \sum_{j,k=1}^{D} \left(h_{ijk}(X^f_n)(\delta w_{j,n} \delta w_{k,n} - \tfrac12 \Omega_{jk} \Delta t) + h_{ijk}(X^a_n)(\delta w_{j,n+\frac12} \delta w_{k,n+\frac12} - \tfrac12 \Omega_{jk} \Delta t) \right). \end{aligned}$$

Following the same method of analysis as in the proof of Lemma 4.7 it can be proved that $\mathbb{E}[\,|N_{i,n}|^p]$ has an $O(\Delta t^p)$ bound.

Next, defining $X^c_{n+\frac12}$ to be the linear interpolant value $\frac12 (X^c_n + X^c_{n+1})$, then the equation for X^c_{n+1} yields

$$X^c_{i,n+\frac12} = X^c_{i,n} + \tfrac12 f_i(X^c_n) \, \Delta t + \tfrac12 \sum_{j=1}^{d} g_{ij}(X^c_n) \, \Delta w_{j,n} + \tfrac12 \sum_{j,k=1}^{d} h_{ijk}(X^c_n)(\Delta w_{j,n} \Delta w_{k,n} - \Omega_{jk} \Delta t). \tag{A.4}$$

Subtracting (A.4) from (A.3) gives

$$\begin{aligned}\overline{X}^f_{i,n+\frac{1}{2}} - X^c_{i,n+\frac{1}{2}} \;\; &= \;\; \overline{X}^f_{i,n} - X^c_{i,n} \;\; + \;\; \tfrac{1}{2}\left(f_i(\overline{X}^f_n) - f_i(X^c_n)\right)\Delta t \\ &\quad + \tfrac{1}{2}\sum_{j=1}^{d}\left(g_{ij}(\overline{X}^f_n) - g_{ij}(X^c_n)\right)\Delta w_{j,n} \\ &\quad + N_{i,n} + \tfrac{1}{2}\sum_{j,k=1}^{d} h_{ijk}(X^c_n)\,(\Delta w_{j,n}\Delta w_{k,n} - \Omega_{jk}\Delta t).\end{aligned}$$

Using the bounds on $\mathbb{E}[\|\overline{X}^f_n - X^c_n\|^p]$, the bounded first derivatives of $f(x)$ and $g(x)$, the uniform bound on $\mathbb{E}[\,|h_{ijk}(X^c_n|^p]$ and standard results for Brownian increments, we can conclude that there exists a constant K_p, independent of both Δt and n, such that such that

$$\mathbb{E}\left[\left\|\overline{X}^f_{i,n+\frac{1}{2}} - X^c_{i,n+\frac{1}{2}}\right\|^p\right] \leq K_p\,\Delta t^p.$$

References

[1] D.L. Burkholder, B. Davis, and R.F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proc. Sixth Berkeley Symposium Math. Statist. Prob., Vol II*, pages 223–240. University of California Press, Berkeley, 1972.

[2] J.M.C. Clark and R.J. Cameron. The maximum rate of convergence of discrete approximations for stochastic differential equations. In B. Grigelionis, editor, *Stochastic Differential Equations, no. 25 in Lecture Notes in Control and Information Sciences*. Springer-Verlag, 1980.

[3] D. Duffie and P. Glynn. Efficient Monte Carlo simulation of security prices. *Annals of Applied Probability*, 5(4):897–905, 1995.

[4] J.G. Gaines and T.J. Lyons. Random generation of stochastic integrals. *SIAM Journal of Applied Mathematics*, 54(4):1132–1146, 1994.

[5] M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer-Verlag, 2008.

[6] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[7] M.B. Giles, D.J. Higham, and X. Mao. Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.

[8] M.B. Giles and C. Reisinger. Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance. Technical report, Mathematical Institute, University of Oxford, 2011.

[9] M.B. Giles and B.J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. In *Advanced Financial Modelling*, Radon Series on Computational and Applied Mathematics, pages 165–181. de Gruyter, 2009.

[10] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.

[11] I. Karatzas and S.E. Shreve. *Broanian Motion and Stochastic Calculus*. Graduate Texts in Mathematics, Vol. 113. Springer, New York, 1991.

[12] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

[13] G.N. Milstein and M.V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer-Verlag, 2004.

[14] T. Müller-Gronbach. *Strong approximation of systems of stochastic differential equations*. Habilitation thesis, TU Darmstadt, 2002.

[15] T. Rydén and M. Wiktorsson. On the simulation of iterated Itô integrals. *Stochastic Processes and their Applications*, 91(1):151–168, 2001.

[16] M. Wiktorsson. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Annals of Applied Probability*, 11(2):470–487, 2001.